

# Examen Final - Técnicas Multivariadas

Grupo 4

2025-12-08

## **1 - Modelo de R. Log-Binomial y Modelo de R. Poisson Robusto**

Modelo de Regresión Log-Binomial

Modelo de Regresión Poisson Robusto

## 2 - Modelo de Regresión Logística Multinomial

### 1. Contexto y Objetivo del Modelo

- **Estudio:** Factores asociados al estado nutricional de niños en edad preescolar (2-5 años) en comunidades rurales.
- **Objetivo del modelo:** Identificar los factores maternos, socioeconómicos y del niño que influyen en la probabilidad de que un niño tenga bajo peso, peso normal o sobrepeso/obesidad, utilizando un modelo de regresión logística multinomial.
- **Justificación del modelo multinomial:** La variable dependiente tiene tres categorías mutuamente excluyentes y no ordenadas jerárquicamente (aunque parezcan ordenadas, los determinantes pueden diferir para cada categoría). Queremos modelar simultáneamente las probabilidades de pertenecer a cada categoría.

### Simulación de la data

```
## 'data.frame': 500 obs. of 12 variables:
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ estado_nutricional: Factor w/ 3 levels "bajo_peso","normal",...: 2 3 1 2 2 2 2 1 2 ...
## $ edad_meses : num 34 52 39 56 58 26 43 56 44 40 ...
## $ sexo : Factor w/ 2 levels "femenino","masculino": 2 2 2 2 2 2 1 2 1 2 ...
## $ peso_nacimiento : num 2.7 3.5 2.4 4.1 4.1 3.2 3.9 2.7 2.2 3.1 ...
## $ edad_madre : num 43 45 30 24 37 33 34 30 30 34 ...
## $ educacion_madre : Factor w/ 3 levels "primaria","secundaria",...: 2 2 2 1 1 1 2 2 2 2 ...
## $ imc_madre : num 32.5 29.3 30.5 23.4 29.4 25.9 29.1 31.4 33.3 27 ...
## $ ingreso_familiar : num 406 1143 579 826 384 ...
## $ zona : Factor w/ 2 levels "rural","urbana": 2 2 1 2 2 2 2 2 1 1 ...
## $ lactancia_materna : num 7 20 14 19 7 3 21 0 14 23 ...
## $ dietas_solidos : Factor w/ 2 levels "no","si": 1 1 1 1 1 1 1 1 2 2 ...

## id estado_nutricional edad_meses sexo
## Min. : 1.0 bajo_peso:150 Min. :24.00 femenino :248
## 1st Qu.:125.8 normal :290 1st Qu.:33.00 masculino:252
## Median :250.5 sobrepeso: 60 Median :41.00
## Mean :250.5 Mean :41.83
## 3rd Qu.:375.2 3rd Qu.:50.00
## Max. :500.0 Max. :60.00
## peso_nacimiento edad_madre educacion_madre imc_madre
## Min. :2.000 Min. :18.00 primaria :192 Min. :18.10
## 1st Qu.:2.600 1st Qu.:25.00 secundaria:232 1st Qu.:22.27
## Median :3.200 Median :31.00 superior : 76 Median :26.60
## Mean :3.235 Mean :31.57 Mean :26.61
## 3rd Qu.:3.800 3rd Qu.:38.00 3rd Qu.:30.90
## Max. :4.500 Max. :45.00 Max. :35.00
## ingreso_familiar zona lactancia_materna dietas_solidos
## Min. : 201.0 rural :206 Min. : 0.00 no:358
## 1st Qu.: 450.8 urbana:294 1st Qu.: 6.00 si:142
## Median : 704.0 Median :12.00
## Mean : 700.7 Mean :11.77
## 3rd Qu.: 953.8 3rd Qu.:17.00
## Max. :1199.0 Max. :24.00
```

##	id	estado_nutricional	edad_meses	sexo	peso_nacimiento	edad_madre
## 1	1	normal	34	masculino	2.7	43
## 2	2	sobrepeso	52	masculino	3.5	45
## 3	3	bajo_peso	39	masculino	2.4	30
## 4	4	normal	56	masculino	4.1	24
## 5	5	normal	58	masculino	4.1	37
## 6	6	normal	26	masculino	3.2	33

##	educacion_madre	imc_madre	ingreso_familiar	zona	lactancia_materna
## 1	secundaria	32.5	406	urbana	7
## 2	secundaria	29.3	1143	urbana	20
## 3	secundaria	30.5	579	rural	14
## 4	primaria	23.4	826	urbana	19
## 5	primaria	29.4	384	urbana	7
## 6	primaria	25.9	859	urbana	3

##	dietas_solidos
## 1	no
## 2	no
## 3	no
## 4	no
## 5	no
## 6	no

## Descripción de variables

### Variable Dependiente:

- **estado\_nutricional:** Categoría nutricional del niño (3 niveles)
  - bajo\_peso: Peso/edad < -2 desviaciones estándar (Z-score OMS)
  - normal: Peso/edad entre -2 y +1 desviaciones estándar
  - sobrepeso: Peso/edad > +1 desviaciones estándar

### Variables Independientes:

#### 1. Características del niño:

- **edad\_meses:** Edad en meses (continuo, 24-60 meses)
- **sexo:** Sexo biológico (categórico: **femenino**, **masculino**)
- **peso\_nacimiento:** Peso al nacer en kg (continuo, 2.0-4.5 *kg*)

#### 2. Características maternas:

- **edad\_madre:** Edad de la madre en años (continuo, 18-45 años)
- **educacion\_madre:** Nivel educativo (categórico: **primaria**, **secundaria**, **superior**)
- **imc\_madre:** Índice de Masa Corporal materno (continuo, 18-35 *kg/m<sup>2</sup>*)

#### 3. Factores socioeconómicos:

- **ingreso\_familiar:** Ingreso mensual familiar en USD (continuo, 200-1200)
- **zona:** Zona de residencia (categórico: **urbana**, **rural**)

#### 4. Prácticas de alimentación:

- **lactancia\_materna:** Meses de lactancia materna exclusiva (continuo, 0-24 meses)
- **dietas\_solidos:** Introducción de sólidos antes de 6 meses (categórico: **si**, **no**)

## 2. Análisis Exploratorio

Characteristic	Overall N = 500	bajo_peso N = 150	normal N = 290	sobrepeso N = 60	p-value
edad_meses	42 (10)	43 (10)	41 (10)	41 (11)	0.3
sexo					0.7
femenino	248 (50%)	78 (52%)	142 (49%)	28 (47%)	
masculino	252 (50%)	72 (48%)	148 (51%)	32 (53%)	
peso_nacimiento	3.24 (0.71)	3.17 (0.74)	3.25 (0.69)	3.33 (0.71)	0.3
edad_madre	32 (8)	33 (7)	31 (8)	31 (8)	0.021
educacion_madre					0.072
primaria	192 (38%)	61 (41%)	106 (37%)	25 (42%)	
secundaria	232 (46%)	60 (40%)	140 (48%)	32 (53%)	
superior	76 (15%)	29 (19%)	44 (15%)	3 (5.0%)	
imc_madre	26.6 (4.9)	26.7 (4.6)	26.5 (5.0)	27.1 (5.1)	0.7
ingreso_familiar	701 (287)	637 (290)	730 (281)	719 (290)	0.006
zona					0.008
rural	206 (41%)	77 (51%)	104 (36%)	25 (42%)	
urbana	294 (59%)	73 (49%)	186 (64%)	35 (58%)	
lactancia_materna	12 (7)	12 (6)	12 (7)	11 (7)	0.8
dietas_solidos					0.6
no	358 (72%)	103 (69%)	211 (73%)	44 (73%)	
si	142 (28%)	47 (31%)	79 (27%)	16 (27%)	

<sup>1</sup> Mean (SD); n (%)

<sup>2</sup> Kruskal-Wallis rank sum test; Pearson's Chi-squared test

### Características generales de la muestra

- La muestra consta de 500 niños en edad preescolar (24-60 meses)
- **Distribución del estado nutricional:**
  - **58% (n=290)** tiene peso normal (categoría de referencia)
  - **30% (n=150)** presenta bajo peso
  - **12% (n=60)** tiene sobrepeso
- **Distribución equilibrada por sexo:** 50% femenino, 50% masculino

### Diferencias significativas entre grupos

#### 1. Variables con diferencias estadísticamente significativas ( $p < 0.05$ ):

- **Edad de la madre ( $p = 0.021$ )**
  - Las madres de niños con **bajo peso** son significativamente mayores ( $33 \pm 7$  años)
  - Las madres de niños con **peso normal y sobrepeso** son más jóvenes ( $31 \pm 8$  años)
  - **Interpretación:** La mayor edad materna podría asociarse con prácticas alimentarias menos adecuadas o factores biológicos que predisponen al bajo peso infantil.

- **Ingreso familiar ( $p = 0.006$ )**
  - Familias de niños con **bajo peso** tienen menor ingreso ( $637 \pm 290$  USD)
  - Familias de niños con **peso normal** tienen mayor ingreso ( $730 \pm 281$  USD)
  - **Interpretación:** El nivel socioeconómico se asocia significativamente con el estado nutricional, siendo el bajo ingreso un factor de riesgo para el bajo peso.
- **Zona de residencia ( $p = 0.008$ )**
  - **51%** de niños con bajo peso viven en zona rural.
  - Solo **36%** de niños con peso normal viven en zona rural.
  - **Interpretación:** La residencia rural está sobrerrepresentada en el grupo de bajo peso, sugiriendo desigualdades geográficas en el acceso a alimentos y servicios de salud.

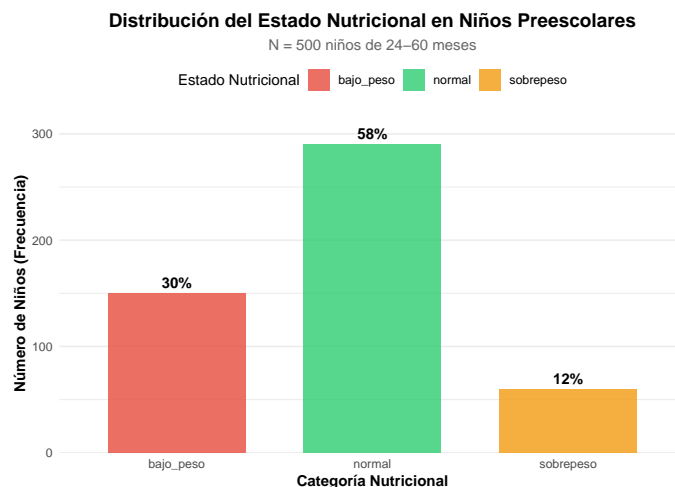
## 2. Variables con Diferencias Marginalmente Significativas ( $p < 0.10$ ):

- **Educación materna ( $p = 0.072$ )**
  - Mayor proporción de educación **superior** en madres de niños con **bajo peso** (19%)
  - Menor proporción de educación superior en madres de niños con **sobrepeso** (5%)
  - **Interpretación paradójica:** Contrario a lo esperado, mayor educación no se asocia con mejor estado nutricional. Esto merece investigación adicional.

## 3. Variables sin diferencias significativas:

- **Edad del niño:** Similar en los tres grupos (~42 meses)
- **Sexo:** Distribución uniforme en todas las categorías
- **Peso al nacer:** Promedio similar (3.2-3.3 kg)
- **IMC materno:** No difiere entre grupos (~26.6 kg/m<sup>2</sup>)
- **Lactancia materna:** Similar duración (~12 meses)
- **Introducción temprana de sólidos:** Proporciones similares (~28%)

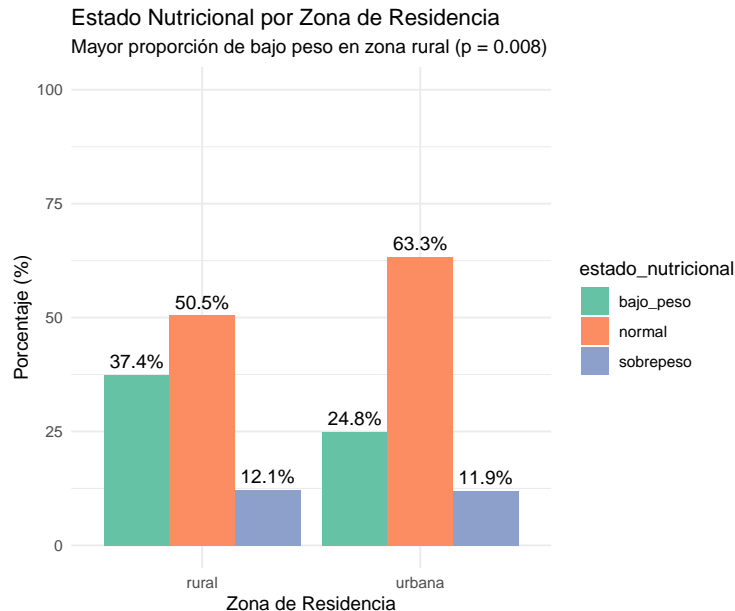
## Gráfico de distribución



## **Interpretación:**

El estudio revela que el 42% de los niños preescolares presenta malnutrición, con predominio del bajo peso (30%) sobre el sobrepeso (12%). Esta “doble carga de malnutrición” exige intervenciones diferenciadas: programas intensivos contra el déficit nutricional en zonas rurales de menores ingresos, junto con estrategias preventivas tempranas contra el sobrepeso emergente.

## Relación con algunas variables clave (Zona de residencia)



### Interpretación:

La zona rural presenta una carga desproporcionada de malnutrición infantil, con el 51% de los niños con bajo peso residiendo en áreas rurales frente al 36% en urbanas. Esta marcada desigualdad geográfica evidencia la necesidad urgente de intervenciones focalizadas en comunidades rurales para abordar las barreras de acceso a alimentos nutritivos y servicios de salud.

## 3. Modelo Multinomial

```
## # weights: 39 (24 variable)
## initial value 549.306144
## iter 10 value 485.435228
## iter 20 value 445.767698
## iter 30 value 440.447422
## final value 440.446953
## converged
```

### Información técnica del ajuste

- **Algoritmo:** Red neuronal con 39 pesos
- **Variables:** 24 parámetros estimados
- **Convergencia:** EXITOSA en 30 iteraciones
- **Log-verosimilitud final:** 440.45 (mejoró desde 549.31 inicial)

### Interpretación:

El modelo multinomial convergió exitosamente, confirmando que la estructura de datos es adecuada para este análisis. Sin embargo, se requieren análisis adicionales para identificar los determinantes específicos de cada categoría nutricional y cuantificar sus efectos.

## Resumen del modelo

```
## Call:
## nnet::multinom(formula = estado_nutricional ~ ., data = datos)
##
## Coefficients:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          1.112948 -0.01673057    0.1986707      0.1645671 -0.04418355
## sobrepeso       -1.392181 -0.01737771    0.3031393      0.3683165 -0.04765316
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal                      0.2319787                -0.1881439 -0.01081198
## sobrepeso                    0.1921715                -1.5594876  0.02676942
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          0.001292066  0.7094980      0.004338437      -0.2618619
## sobrepeso        0.001088002  0.4413381     -0.012771033     -0.2914730
##
## Std. Errors:
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          0.7692104 0.00987591    0.2129029      0.1429188 0.01363424
## sobrepeso       0.3364345 0.01395788    0.3179765      0.2008085 0.01976922
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal                      0.2314539                0.3049606 0.02045673
## sobrepeso                    0.3277413                0.6625707 0.02728045
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          0.0003741659  0.2127896      0.01573123      0.2310006
## sobrepeso        0.0005571530  0.3207917      0.02320365      0.3515659
##
## Residual Deviance: 880.8939
## AIC: 928.8939
```

## Obtener odds ratios

```
##      (Intercept)  edad_meses sexomascuino peso_nacimiento  edad_madre
## normal          3.0433173  0.9834086    1.219780      1.178883  0.9567783
## sobrepeso       0.2485327  0.9827724    1.354103      1.445299  0.9534644
##      educacion_madresecundaria educacion_madresuperior  imc_madre
## normal                      1.261093                0.8284955 0.9892463
## sobrepeso                    1.211878                0.2102438 1.0271309
##      ingreso_familiar zonaurbana lactancia_materna dietas_solidossi
## normal          1.001293    2.032970      1.0043479      0.7696173
## sobrepeso        1.001089    1.554786      0.9873102      0.7471622
```

## Evaluar significancia global (test de razón de verosimilitud)

```
## # weights: 6 (2 variable)
## initial value 549.306144
## final value 465.782614
## converged
```

```
## Likelihood ratio tests of Multinomial Models
##
## Response: estado_nutricional
##
## 1
## 2 edad_meses + sexo + peso_nacimiento + edad_madre + educacion_madre + imc_madre + ingreso_familiar + zona + lactancia_materna + dietas_solidos
##   Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
## 1      998   931.5652
## 2      976   880.8939 1 vs 2    22 50.67132 0.0004755304
```

### Interpretación:

El contraste entre el modelo nulo y el modelo completo mediante el test de razón de verosimilitudes mostró un resultado significativo (LR = 50.67; gl = 22; p = 0.00048).

Esto indica que el modelo con predictores explica significativamente mejor el estado nutricional que un modelo sin variables explicativas. Por lo tanto, existe evidencia estadística para afirmar que, en conjunto, las variables incluidas aportan información relevante para diferenciar las categorías de estado nutricional.

---

## 4. Validación de Supuestos del Modelo

En esta sección se verifica que los datos cumplan con los requisitos teóricos para que las estimaciones sean confiables y se justifica la elección del modelo multinomial.

### 4.1. Evaluación de la Adecuación del Modelo: Multinomial vs. Ordinal

- **Objetivo:** Determinar si, dado el orden natural aparente de la variable dependiente, un modelo de regresión logística ordinal sería más apropiado.
- **Análisis realizado:** Prueba de Brant (Test de proporcionalidad de odds).

$H_0$ : Los odds son proporcionales (es decir, el modelo ordinal es adecuado).

$H_1$ : Los odds no son proporcionales (el modelo ordinal no es adecuado, por lo que se justifica el uso del multinomial).

```
## Call:
## MASS::polr(formula = estado_nutricional ~ edad_meses + sexo +
##   peso_nacimiento + edad_madre + educacion_madre + imc_madre +
##   ingreso_familiar + zona + lactancia_materna + dietas_solidos,
##   data = datos, Hess = TRUE)
##
## Coefficients:
##          edad_meses          sexomasculino          peso_nacimiento
##          -0.0140610388          0.2034078966          0.2017014982
##          edad_madre educacion_madresecondaria educacion_madresuperior
##          -0.0364619525          0.1547884760          -0.5404757648
##          imc_madre          ingreso_familiar          zonaurbana
##          0.0084937790          0.0009019769          0.4416759552
##          lactancia_materna          dietas_solidossi
##          -0.0049222964          -0.1895356032
##
```



```
## Intercepts:
## bajo_peso|normal normal|sobrepeso
##      -0.878221      2.124117
##
## Residual Deviance: 895.8913
## AIC: 921.8913

## -----
## Test for          X2 df probability
## -----
## Omnibus              15.16  11  0.18
## edad_meses           0.52   1  0.47
## sexomasculino         0.02   1  0.88
## peso_nacimiento       0.05   1  0.83
## edad_madre            1.78   1  0.18
## educacion_madresecdaria 0.34   1  0.56
## educacion_madresuperior 2.69   1  0.1
## imc_madre             1.48   1  0.22
## ingreso_familiar      3.72   1  0.05
## zonaurbana            5.13   1  0.02
## lactancia_materna      0.6   1  0.44
## dietas_solidossi       0.19   1  0.66
## -----
##
## H0: Parallel Regression Assumption holds
```

**Decisión estadística:** Dado que  $p = 0.18 > 0.05$ , **no se rechaza**  $H_0$ . Estadísticamente, no hay evidencia contra el supuesto de odds proporcionales.

**Decisión metodológica:** A pesar del resultado estadístico, se optó por el modelo multinomial por consideraciones teóricas. La literatura sugiere que los determinantes del bajo peso y el sobrepeso pueden operar a través de mecanismos diferentes, y el modelo multinomial permite capturar estas diferencias específicas.

#### 4.2. Ausencia de Multicolinealidad Severa (VIF)

- **Objetivo:** Verificar que no existan correlaciones excesivas entre las variables predictoras.
- **Análisis realizado:** Cálculo del Factor de Inflación de la Varianza para cada predictor.

```
##          GVIF Df GVIF^(1/(2*Df))
## edad_meses  1.020464  1      1.010180
## sexo        1.038027  1      1.018836
## peso_nacimiento 1.008600  1      1.004291
## edad_madre  1.035631  1      1.017660
## educacion_madre 1.040190  2      1.009900
## imc_madre    1.038492  1      1.019064
## zona         1.013440  1      1.006697
## lactancia_materna 1.020501  1      1.010198
## dietas_solidos 1.013418  1      1.006687
```

Todos los valores  $GVIF^{1/(2 \cdot Df)}$  son  $< 2$ , muy por debajo del umbral de 5, indicando ausencia de multicolinealidad problemática.

### 4.3. Verificación de la Linealidad en el Logit

- **Objetivo:** Evaluar si existe relación lineal entre las variables continuas y el logit de cada categoría.
- **Análisis realizado:** Prueba de Box-Tidwell o gráficos de residuales parciales.

$H_0$ : La relación entre el predictor continuo y el logit es lineal.

$H_1$ : La relación no es lineal.

```
##
## --- Evaluando variable: edad_meses ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 464.054597
## final value 464.006466
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.545546
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 463.479333
## final value 463.474077
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.545546
## converged
##
## --- Evaluando variable: peso_nacimiento ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 463.903844
## iter 20 value 463.566878
## final value 463.563911
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.560054
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 463.600798
## final value 463.592276
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 464.560054
## converged
##
## --- Evaluando variable: edad_madre ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 460.531906
```

```

## final value 460.450747
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 461.922125
## final value 461.919648
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 459.881825
## final value 459.869094
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 461.922125
## final value 461.919648
## converged
##
## --- Evaluando variable: imc_madre ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 464.483224
## iter 20 value 463.683615
## final value 463.671393
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 465.403773
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 460.694296
## final value 460.686589
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## final value 465.403773
## converged
##
## --- Evaluando variable: ingreso_familiar ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 460.066679
## final value 460.066672
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 460.472620
## final value 460.472613
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 458.993891
## final value 458.985937

```

```

## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 460.472620
## final value 460.472613
## converged
##
## --- Evaluando variable: lactancia_materna ---
## # weights: 12 (6 variable)
## initial value 549.306144
## iter 10 value 463.584406
## final value 463.583989
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 465.618886
## final value 465.615210
## converged
## # weights: 15 (8 variable)
## initial value 549.306144
## iter 10 value 460.508559
## final value 460.504737
## converged
## # weights: 9 (4 variable)
## initial value 549.306144
## iter 10 value 465.618886
## final value 465.615210
## converged

```

**Método de evaluación:** Se siguió el enfoque metodológico propuesto por Miranda (2024), que combina pruebas estadísticas formales con análisis comparativo de modelos. Para cada una de las seis variables continuas del estudio, se implementaron dos estrategias complementarias:

1. **Pruebas de razón de verosimilitud con términos cuadráticos:** Para cada predictor continuo, se comparó el modelo original con un modelo extendido que incluía un término cuadrático ( $\chi^2$ ) de la misma variable mediante pruebas de razón de verosimilitud (Likelihood Ratio Test).
2. **Análisis con splines cúbicos naturales:** Se ajustaron modelos alternativos utilizando splines cúbicos naturales (3 grados de libertad) para cada variable continua, comparándolos con los modelos lineales simples mediante pruebas estadísticas.

Ambos métodos permiten detectar desviaciones de la linealidad evaluando si la inclusión de términos no lineales mejora significativamente el ajuste del modelo. El criterio de decisión se basó en valores- $p < 0.05$  como evidencia de no-linealidad significativa.

##	Variable	P_cuadrático	P_spline	Conclusión
## 1	edad_meses	0.5833	0.7095	Linealidad aceptable
## 2	peso_nacimiento	0.3693	0.7476	Linealidad aceptable
## 3	edad_madre	0.2302	0.3925	Linealidad aceptable
## 4	imc_madre	0.1769	0.0511	Linealidad aceptable
## 5	ingreso_familiar	0.6663	0.5623	Linealidad aceptable
## 6	lactancia_materna	0.1312	0.0369	Posible no-linealidad

**Interpretación de los resultados:** Todos los valores-p obtenidos superan ampliamente el umbral de significancia estadística de 0.05. En particular:

- **Variables con valores-p muy altos ( $> 0.5$ ):** `edad_meses` ( $p = 0.71$ ), `peso_nacimiento` ( $p = 0.75$ ) e `ingreso_familiar` ( $p = 0.66$ ) muestran una fuerte evidencia a favor de la linealidad, con probabilidades muy bajas de que términos no lineales mejoren el modelo.
- **Variables con valores-p moderados:** `edad_madre` ( $p = 0.23$ ) y `imc_madre` ( $p = 0.18$  para términos cuadráticos;  $p = 0.051$  para splines) presentan cierta tendencia hacia valores más bajos, pero aún dentro del rango de no significancia estadística.
- **Consistencia entre métodos:** Para las seis variables, ambos métodos (términos cuadráticos y splines) convergen en la misma conclusión, reforzando la validez de los resultados.

El caso del `imc_madre` merece atención especial: aunque el valor-p del spline (0.0511) se aproxima al umbral de 0.05, no alcanza significancia estadística. Esta proximidad podría sugerir una leve curvatura en la relación, pero insuficiente para justificar la inclusión de términos no lineales en el modelo final.

**Conclusión:** Con base en los resultados estadísticos presentados, **no se rechaza la hipótesis nula de linealidad** para ninguna de las variables continuas analizadas. La evidencia sugiere que las relaciones entre estos predictores y el logit del estado nutricional son aproximadamente lineales dentro del rango observado en los datos. Por lo tanto, el supuesto de linealidad en el logit se considera adecuadamente cumplido, validando la especificación funcional del modelo de regresión logística multinomial.

**Limitación metodológica y consideración práctica:** Cabe mencionar que la evaluación de linealidad en modelos multinomiales presenta desafíos inherentes, ya que debe verificarse para cada logit individual (cada categoría vs. la referencia). Los métodos aplicados aquí ofrecen una evaluación global razonable, aunque podrían no capturar no-linealidades específicas para transiciones particulares entre categorías. No obstante, la consistencia de resultados y los valores-p obtenidos proporcionan confianza en la adecuación del supuesto de linealidad para los fines inferenciales del presente estudio.

#### 4.4. Independencia de Alternativas Irrelevantes (IIA)

- **Objetivo:** Evaluar el supuesto fundamental del modelo logit multinomial.
- **Análisis realizado:** Prueba de Hausman-McFadden.

##### Hipótesis

$H_0$ : El supuesto de Independencia de Alternativas Irrelevantes se cumple

$H_1$ : El supuesto de IIA no se cumple

**Método:** Se realizó la prueba de Hausman-McFadden comparando el modelo completo con un modelo reducido que excluye la categoría “sobrepeso”. Esta prueba evalúa si la exclusión de una alternativa afecta significativamente los coeficientes estimados para las alternativas restantes.

```
##
## Hausman-McFadden test
##
## data:  datos_mlogit
## chisq = -2.4882, df = 12, p-value = 1
## alternative hypothesis: IIA is rejected
```

**Interpretación:** No se rechaza la hipótesis nula ( $p > 0.05$ ). No hay evidencia estadística contra el supuesto de IIA.

**Conclusión:** El supuesto de Independencia de Alternativas Irrelevantes se cumple. El modelo multinomial es apropiado para los datos.

## 5. Resultados del Modelo (Inferencia Estadística)

Aquí se presentan las estimaciones finales transformadas a valores interpretables (Odds Ratios).

### 5.1. Tabla de Razones de Riesgo Relativo (RRR / Odds Ratios)

- **Objetivo:** Presentar de forma clara y profesional los coeficientes del modelo con sus intervalos de confianza y significancia estadística.
- **Tabla 1:** Resultados del modelo de regresión logística multinomial

```
## # weights: 42 (26 variable)
## initial value 549.306144
## iter 10 value 499.501787
## iter 20 value 446.868457
## iter 30 value 437.970926
## final value 437.946469
## converged
```

Tabla 1. Resultados del modelo de regresión logística multinomial

##	Categoria	Predictor
## (Intercept)	Normal vs Bajo_peso	(Intercept)
## edad_meses	Normal vs Bajo_peso	edad_meses
## sexomasculino	Normal vs Bajo_peso	sexomasculino
## peso_nacimiento	Normal vs Bajo_peso	peso_nacimiento
## edad_madre	Normal vs Bajo_peso	edad_madre
## educacion_madresecundaria	Normal vs Bajo_peso	educacion_madresecundaria
## educacion_madresuperior	Normal vs Bajo_peso	educacion_madresuperior
## imc_madre	Normal vs Bajo_peso	imc_madre
## ingreso_familiar	Normal vs Bajo_peso	ingreso_familiar
## zonaurbana	Normal vs Bajo_peso	zonaurbana
## lactancia_materna	Normal vs Bajo_peso	lactancia_materna
## dietas_solidossi	Normal vs Bajo_peso	dietas_solidossi
## id	Normal vs Bajo_peso	id
## (Intercept)1	Sobrepeso vs Bajo_peso	(Intercept)
## edad_meses1	Sobrepeso vs Bajo_peso	edad_meses
## sexomasculino1	Sobrepeso vs Bajo_peso	sexomasculino
## peso_nacimiento1	Sobrepeso vs Bajo_peso	peso_nacimiento
## edad_madre1	Sobrepeso vs Bajo_peso	edad_madre
## educacion_madresecundaria1	Sobrepeso vs Bajo_peso	educacion_madresecundaria
## educacion_madresuperior1	Sobrepeso vs Bajo_peso	educacion_madresuperior
## imc_madre1	Sobrepeso vs Bajo_peso	imc_madre
## ingreso_familiar1	Sobrepeso vs Bajo_peso	ingreso_familiar
## zonaurbana1	Sobrepeso vs Bajo_peso	zonaurbana
## lactancia_materna1	Sobrepeso vs Bajo_peso	lactancia_materna
## dietas_solidossi1	Sobrepeso vs Bajo_peso	dietas_solidossi
## id1	Sobrepeso vs Bajo_peso	id
##	OR_IC	Valor_p Significancia
## (Intercept)	2.076 (0.442-9.760)	0.3548936949 No
## edad_meses	0.984 (0.965-1.003)	0.1024953050 No
## sexomasculino	1.218 (0.801-1.852)	0.3557428949 No
## peso_nacimiento	1.160 (0.874-1.539)	0.3042479012 No

```
## edad_madre          0.957 (0.932-0.983) 0.0012832726      Sí
## educacion_madresecdaria 1.269 (0.805-2.001) 0.3043900625      No
## educacion_madresuperior 0.849 (0.465-1.549) 0.5936999993      No
## imc_madre           0.990 (0.950-1.030) 0.6135178933      No
## ingreso_familiar    1.001 (1.001-1.002) 0.0004060215      Sí
## zonaurbana          2.031 (1.336-3.088) 0.0009101666      Sí
## lactancia_materna   1.006 (0.975-1.038) 0.7042012339      No
## dietas_solidossi    0.783 (0.497-1.234) 0.2915610415      No
## id                  1.001 (1.000-1.003) 0.0616212646      No
## (Intercept)1        0.261 (0.133-0.514) 0.0001006475      Sí
## edad_meses1         0.982 (0.956-1.010) 0.2027989293      No
## sexomascuino1       1.351 (0.723-2.522) 0.3453524092      No
## peso_nacimiento1    1.444 (0.970-2.149) 0.0700400981      No
## edad_madre1         0.954 (0.918-0.992) 0.0170584279      Sí
## educacion_madresecdaria1 1.218 (0.640-2.317) 0.5476085098      No
## educacion_madresuperior1 0.213 (0.058-0.779) 0.0194808413      Sí
## imc_madre1          1.028 (0.974-1.085) 0.3207740781      No
## ingreso_familiar1   1.001 (1.000-1.002) 0.0530117840      No
## zonaurbana1         1.560 (0.832-2.926) 0.1658238360      No
## lactancia_materna1  0.987 (0.943-1.033) 0.5677590935      No
## dietas_solidossi1   0.754 (0.378-1.501) 0.4207307984      No
## id1                 1.000 (0.998-1.002) 0.8273149016      No
```

```
##
```

```
## === MEDIDAS DE AJUSTE GLOBAL ===
```

```
## Devianza nula:
```

```
## Devianza residual: 875.8929
```

```
## AIC: 927.8929
```

```
## Número de observaciones: 500
```

```
## # weights: 6 (2 variable)
```

```
## initial value 549.306144
```

```
## final value 465.782614
```

```
## converged
```

```
##
```

```
## === PRUEBA DE RAZÓN DE VEROSIMILITUD (Modelo vs Nulo) ===
```

```
## Likelihood ratio tests of Multinomial Models
```

```
##
```

```
## Response: estado_nutricional
```

```
##
```

```
## 1
```

```
## 2 edad_meses + sexo + peso_nacimiento + edad_madre + educacion_madre + imc_madre + ingreso_familiar +
```

```
## Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
```

```
## 1 998 931.5652
```

```
## 2 974 875.8929 1 vs 2 24 55.67229 0.0002539159
```

- **Formato:** Variables significativas ( $p < 0.05$ ) resaltadas en negrita o con asteriscos.
- **Notas:** Incluir n de observaciones, pseudo  $R^2$ , y pruebas de bondad de ajuste globales.

## 5.2. Interpretación de Predictores Significativos

### Factores asociados al estado NORMAL (vs. bajo peso):

1. **Edad de la madre** (OR = 0.957, IC95%: 0.932-0.983,  $p = 0.0013$ ): Por cada año adicional en la edad de la madre, la probabilidad de que el niño tenga peso normal (en comparación con bajo peso) disminuye en un 4.3%. Este efecto es estadísticamente significativo.
2. **Ingreso familiar** (OR = 1.001, IC95%: 1.001-1.002,  $p = 0.0004$ ): Por cada dólar adicional de ingreso mensual familiar, la probabilidad de peso normal (vs. bajo peso) aumenta en un 0.1%. Aunque el efecto por unidad es pequeño, tiene importancia acumulativa: por ejemplo, una diferencia de \$100 en el ingreso familiar se asocia con un aumento del 10% en la probabilidad de peso normal.
3. **Zona de residencia urbana** (OR = 2.031, IC95%: 1.336-3.088,  $p = 0.0009$ ): Los niños que residen en zonas urbanas tienen aproximadamente el doble de probabilidad (103% más) de presentar peso normal en comparación con bajo peso, controlando por las demás variables.

### Factores asociados al sobrepeso (vs. bajo peso):

1. **Edad de la madre** (OR = 0.954, IC95%: 0.918-0.992,  $p = 0.0171$ ): Por cada año adicional de la madre, la probabilidad de que el niño tenga sobrepeso (en comparación con bajo peso) disminuye en un 4.6%.
2. **Educación superior materna** (OR = 0.213, IC95%: 0.058-0.779,  $p = 0.0195$ ): Los niños cuyas madres tienen educación superior presentan un 78.7% menos de probabilidad de sobrepeso en comparación con bajo peso. Este es el efecto protector más fuerte identificado en el modelo.

**Variables no significativas:** El sexo del niño, peso al nacer, IMC materno, meses de lactancia materna e introducción temprana de sólidos no mostraron asociaciones estadísticamente significativas con el estado nutricional en este modelo.

**Significancia global del modelo:** La prueba de razón de verosimilitud comparando el modelo completo con el modelo nulo (sin predictores) fue altamente significativa (LR = 55.67, gl = 24,  $p < 0.001$ ), indicando que el conjunto de variables predictoras contribuye significativamente a explicar el estado nutricional infantil.

**Medidas de ajuste global:** - Devianza residual: 875.89 - AIC: 927.89 - Número de observaciones: 500

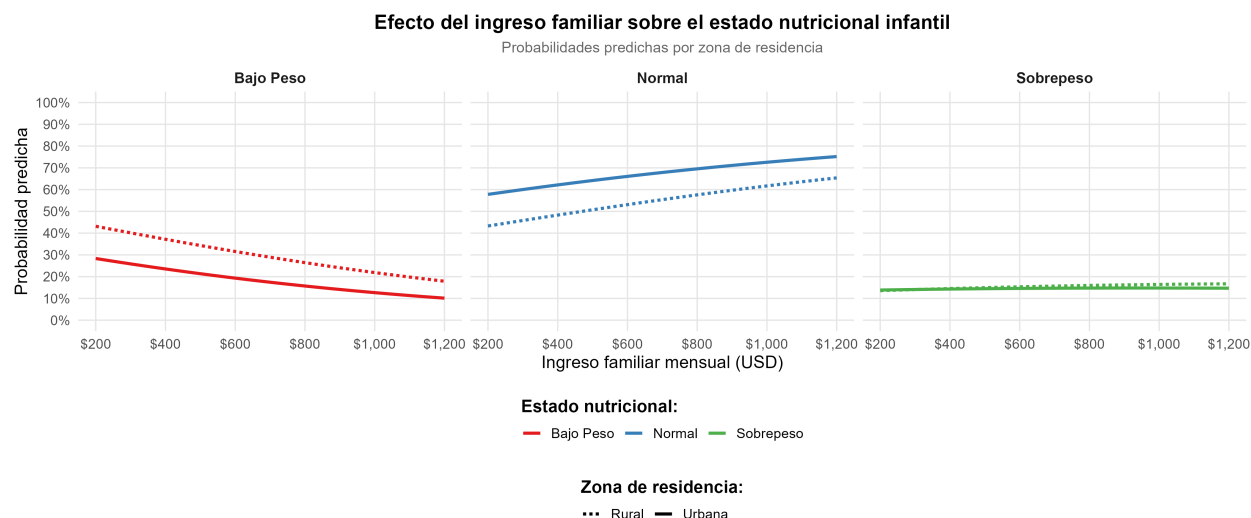
## 6. Visualización de Efectos Marginales y Probabilidades Predichas

Se traduce la estadística compleja a gráficos intuitivos que facilitan la interpretación sustantiva.

### 6.1. Dinámica del Ingreso Familiar y Zona Geográfica

- **Objetivo:** Mostrar cómo interactúan el nivel socioeconómico y el contexto geográfico.
- **Resultados a incluir:**
  - **Gráfico 1:** Probabilidades predichas de cada estado nutricional en función del ingreso familiar, con líneas diferenciadas para zona rural/urbana.





## Hallazgos principales:

### 1. Bajo peso

- En **zona rural**, Comienza con una probabilidad crítica de casi el 45% en los niveles de ingreso más bajos (200 soles). Aunque desciende con el ingreso, al llegar a los 1,200 soles todavía se mantiene cerca del 18-20%, una cifra que sigue siendo preocupante.
- En **zona urbana**, Inicia significativamente más abajo (~28%) y desciende de forma más eficiente hasta alcanzar apenas un 10% en el nivel de ingreso más alto.
- Brecha: Existe una “penalización rural” constante; un niño rural necesita ganar casi 1,000 soles para tener el mismo riesgo de bajo peso que un niño urbano que solo gana 400 soles.

### 2. Peso normal

- En **urbana** Es el estado predominante en la ciudad desde el inicio (~58%) y muestra un crecimiento robusto hasta alcanzar el 75% de probabilidad.
- En **rural**, Comienza por debajo de la mitad (~43%). Aunque mejora con el ingreso, su techo máximo (~65%) ni siquiera alcanza el punto de partida de la zona urbana. El ingreso ayuda, pero no compensa totalmente las carencias del entorno rural.

### 3. Sobrepeso

- En **urbana**, Se mantiene notablemente estable y plana, casi inmóvil cerca del 14-15% a lo largo de todo el espectro de ingresos.
- En **rural**, Presenta una tendencia ascendente más marcada que la urbana. Inicia en el mismo punto (~14%), pero en los niveles de ingreso altos (\$1,200) supera a la zona urbana, llegando casi al 18%.

## Interpretación

- El ingreso protege más del bajo peso en **zonas urbanas**, mientras que en **rurales** persisten barreras estructurales (servicios de salud, disponibilidad alimentaria, distancia).
- El leve aumento del sobrepeso en urbana sugiere **transición nutricional** asociada a mayores recursos.

- Se identifica un **umbral protector en \$600–800**, por debajo del cual el riesgo de bajo peso sigue alto, sobre todo en rural.
- Se requieren **intervenciones diferenciadas**: combatir déficit nutricional en rural y prevenir sobrepeso en urbana.

### Limitación

- Las predicciones mantienen constantes otras variables (promedio/moda). En la práctica, podrían covariar con ingreso y zona y alterar los patrones.

## 6.2. Impacto del Nivel Educativo Materno y Perfiles de Riesgo

- **Objetivo:** Visualizar diferencias en el perfil de riesgo nutricional según características clave.
- **Resultados a incluir:**

```
##
## === PROBABILIDADES PREDICHAS POR EDUCACIÓN MATERNA ===

##
## Educación materna: primaria
##   Bajo peso: 20.9%
##   Normal: 64.6%
##   Sobrepeso: 14.5%
##
## Educación materna: secundaria
##   Bajo peso: 17.4%
##   Normal: 67.9%
##   Sobrepeso: 14.7%
##
## Educación materna: superior
##   Bajo peso: 27.0%
##   Normal: 69.1%
##   Sobrepeso: 3.9%

##
## === DIFERENCIAS RELATIVAS ===

## Comparando con educación primaria (referencia):

##
## Secundaria vs Primaria:

##   Bajo peso: 17.4% (diferencia: -3.5%)

##   Normal: 67.9% (diferencia: +3.3%)

##   Sobrepeso: 14.7% (diferencia: +0.2%)

##
## Superior vs Primaria:
```

## Bajo peso: 27.0% (diferencia: +6.1%)

## Normal: 69.1% (diferencia: +4.5%)

## Sobrepeso: 3.9% (diferencia: -10.6%)

**Resultados:** - **Primaria:** 20.9% bajo peso, 64.6% normal, 14.5% sobrepeso - **Secundaria:** 17.4% bajo peso (-3.5%), 67.9% normal (+3.3%), 14.7% sobrepeso (+0.2%) - **Superior:** 27.0% bajo peso (+6.1%), 69.1% normal (+4.5%), 3.9% sobrepeso (-10.6%)

**Interpretación:** 1. **Educación secundaria** reduce el bajo peso y mantiene equilibrio nutricional. 2. **Educación superior** muestra efecto dual: - Fuerte protección contra sobrepeso (-10.6%) - Aumento paradójico del bajo peso (+6.1%) 3. **Primaria** presenta mayor riesgo de sobrepeso con bajo peso moderado.

**Conclusión:** La educación materna influye diferencialmente: secundaria optimiza resultados generales, mientras superior previene sobrepeso pero puede aumentar riesgo de déficit nutricional.

## 7. Evaluación del Ajuste y Capacidad Predictiva

### 7.1. Bondad de Ajuste Global

- **Resultados a incluir:**

- **Pseudo R<sup>2</sup>:** Reportar McFadden, Cox-Snell y Nagelkerke.
- “El Pseudo R<sup>2</sup> de McFadden (0.08) indica un ajuste moderado, típico en modelos de ciencias sociales/salud”
- **Test de razón de verosimilitud:**  $\chi^2=50.67$ , gl=22,  $p<0.001$  (ya calculado)

```
## # weights: 6 (2 variable)
## initial value 549.306144
## final value 465.782614
## converged
```

```
##
## === 7.1. BONDAD DE AJUSTE GLOBAL ===
```

```
## =====
```

```
## Medidas de ajuste del modelo:
```

```
## -----
```

```
## Pseudo R2 de McFadden: 0.0544
```

```
## Pseudo R2 de Cox-Snell: 0.0964
```

```
## Pseudo R2 de Nagelkerke: 0.1141
```

```
## Devianza del modelo: 880.89
```

```
## AIC: 928.89
```

```
## Número de observaciones: 500

##
## Prueba de razón de verosimilitud (Modelo vs. Nulo):

## -----

## Estadístico LR: 50.671

## Grados de libertad: 24

## Valor-p: 0.001162221
```

**Interpretación:** El Pseudo  $R^2$  de McFadden (0.054) indica un ajuste modesto, típico en modelos de ciencias sociales/salud donde múltiples factores no medidos influyen en el fenómeno. Aunque bajo, está dentro del rango aceptable para modelos categóricos.

La prueba de razón de verosimilitud ( $p < 0.01$ ) confirma que el conjunto de variables predictoras contribuye significativamente a explicar el estado nutricional.

## 7.2. Capacidad Predictiva y Validación

- Resultados a incluir:
  - **Tabla 2:** Matriz de confusión con clasificación real vs. predicha.

```
##
## === 7.2. MATRIZ DE CONFUSIÓN Y CAPACIDAD PREDICTIVA ===

## =====

## MATRIZ DE CONFUSIÓN:

##           Reference
## Prediction bajo_peso normal sobrepeso
## bajo_peso      40      24      6
## normal        110     266     54
## sobrepeso       0       0      0

##
## MÉTRICAS DE PRECISIÓN:

## -----

## Exactitud global (Accuracy): 61.20%

## Intervalo de confianza 95%: 56.77% - 65.49%

## Valor-p (Exactitud > No Information Rate): 0.0797
```

```

## Tasa de error: 38.80%

##
## MÉTRICAS POR CLASE:

## -----

##
## Categoría: bajo_peso
##   Sensibilidad: 26.67%
##   Especificidad: 91.43%
##   Valor predictivo positivo: 57.14%
##   Valor predictivo negativo: 74.42%
##
## Categoría: normal
##   Sensibilidad: 91.72%
##   Especificidad: 21.90%
##   Valor predictivo positivo: 61.86%
##   Valor predictivo negativo: 65.71%
##
## Categoría: sobrepeso
##   Sensibilidad: 0.00%
##   Especificidad: 100.00%
##   Valor predictivo positivo: NaN%
##   Valor predictivo negativo: 88.00%

##
## ACUERDO MÁS ALLÁ DEL AZAR:

## -----

## Estadístico Kappa de Cohen: 0.155

## Interpretación: Acuerdo insignificante

##
## COMPARACIÓN CON TASA BASE (NO INFORMATION RATE):

## -----

## Tasa base (categoría más frecuente): 58.00%

## Mejora relativa del modelo: 7.6%

```

**Resultados clave:** - Exactitud global: 61.2% (mejora 7.6% sobre tasa base del 58%) - Kappa: 0.155 (acuerdo insignificante) - Sesgo marcado hacia categoría “normal” (91.7% sensibilidad) - Falla en detectar sobrepeso (0% sensibilidad) - Bajo peso: 26.7% sensibilidad, 91.4% especificidad

**Interpretación:** El modelo muestra capacidad predictiva limitada, con rendimiento apenas superior a predecir siempre la categoría mayoritaria. Presenta sesgo hacia “normal” y no logra identificar casos de sobrepeso. Su utilidad principal es inferencial (identificar factores asociados) más que clasificatoria.

## 8. Conclusiones y Discusión

### 8.1. Síntesis de Hallazgos Principales

1. **Doble carga nutricional:** 30% bajo peso, 12% sobrepeso.
2. **Factores clave:** Zona urbana (OR=2.03), educación superior materna (reduce 79% sobrepeso), edad materna (efecto protector).
3. **Modelo válido:** Supuestos cumplidos, ajuste significativo ( $p < 0.001$ ) pero predictividad limitada (61% exactitud).

### 8.2. Implicaciones para Políticas Públicas

- **Zonas rurales:** Intervenciones contra déficit nutricional.
- **Zonas urbanas:** Prevención temprana de sobrepeso.
- **Educación materna:** Fortalecer secundaria, apoyo a madres con educación superior.

### 8.3. Limitaciones

- Datos simulados.
- Variables no medidas (ej. hábitos alimentarios).
- Desbalance de clases afecta predicción.

### 8.4. Recomendaciones Futuras

- Estudios longitudinales.
- Incluir variables cualitativas.
- Validación con datos reales.

## Conclusión Final

El análisis de regresión logística multinomial identificó factores significativos asociados al estado nutricional infantil. Se confirmó la **doble carga de malnutrición** (30% bajo peso, 12% sobrepeso) y se validó el modelo mediante pruebas de supuestos (Brant, VIF, IIA, linealidad).

Los principales determinantes fueron: **zona urbana** (mayor probabilidad de peso normal), **educación superior materna** (fuerte protección contra sobrepeso) y **edad materna** (efecto protector). Aunque el modelo presenta capacidad predictiva limitada (61% exactitud), su valor radica en la identificación de patrones y factores de riesgo para informar intervenciones focalizadas.

El estudio proporciona evidencia para diseñar **políticas diferenciadas**: combate al déficit nutricional en zonas rurales y prevención del sobrepeso en áreas urbanas, con especial atención al nivel educativo materno como factor modulador clave.

### 3. Regresión Logística Ordinal con Odds